Raymond Lin
304937942

CS M146 - Problem Set 5

1. $(h_t^*(x), \beta^*_t) = \underset{(h_t(x), \beta_t)}{\text{argmin}} \ (e^{\beta t} - e^{-\beta t}) \sum_n w_t(n) \mathbb{I}(y_n \neq h_t(x_n)) + e^{-\beta t} \sum_n w_t(n)$

$\mathcal{E}_t = \sum_n w_t(n) \mathbb{I}(y_n \neq h_t(x_n))$. s.t. $\sum_n w_t(n) = 1$

a) $\frac{\partial}{\partial \beta t}\left[(e^{\beta t} - e^{-\beta t}) \sum_n w_t(n) \mathbb{I}(y_n \neq h_t(x_n)) + e^{-\beta t} \sum_n w_t(n)\right]$

$= \frac{\partial}{\partial \beta t}\left[(e^{\beta t} - e^{-\beta t}) \mathcal{E}_t + e^{-\beta t}\right]$

$= \frac{\partial}{\partial \beta t}\left(e^{\beta t}\mathcal{E}_t - e^{-\beta t}\mathcal{E}_t + e^{-\beta t}\right)$

$= \beta e^{\beta t}\mathcal{E}_t + \beta e^{-\beta t}\mathcal{E}_t - \beta e^{-\beta t} = 0$

$\cancel{\frac{\beta e^{-\beta t}}{\beta e^{-\beta t}}} = \frac{\beta e^{\beta t}\mathcal{E}_t}{\beta e^{-\beta t}} + \frac{\cancel{\beta e^{-\beta t}}\mathcal{E}_t}{\cancel{\beta e^{-\beta t}}}$

$1 = e^{2\beta t}\mathcal{E}_t + \mathcal{E}_t$

$\frac{1-\mathcal{E}t}{\mathcal{E}t} = e^{2\beta t}$

$\ln\left(\frac{1-\mathcal{E}t}{\mathcal{E}t}\right) = 2\beta t$

$\beta t = \frac{1}{2}\ln\left(\frac{1-\mathcal{E}t}{\mathcal{E}t}\right)$

$$\boxed{\beta t^* = \frac{1}{2}\log\left(\frac{1-\mathcal{E}t}{\mathcal{E}t}\right)}$$

b) Hard margin SVM $\Rightarrow$ $\mathcal{E}_t = 0$

$\beta_1^* = \frac{1}{2}\log\left(\frac{1-\mathcal{E}_1}{\mathcal{E}_1}\right) = \frac{1}{2}\log\left(\frac{1}{0}\right)$

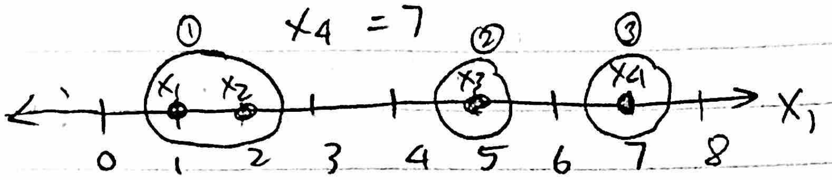
$$\boxed{\beta_1^* = \infty}$$

2\. a) $K=3$, $x_1 = 1$

$x_2 = 2$

$x_3 = 5$

$x_4 = 7$



optimal clustering: $\{x_1, x_2\}$ ①, $\{x_3\}$ ②, $\{x_4\}$ ③

$$J(\{r_{nk}\}, \{\mu_k\}) = \sum_n \sum_k r_{nk} \|x_n - \mu_k\|_2^2 \qquad \mu_1 = \frac{3}{2}$$

$$= 1(1-\tfrac{3}{2})^2 + 1(2-\tfrac{3}{2})^2 + 1(5-5)^2 + 1(7-7)^2 \qquad \mu_2 = 5$$

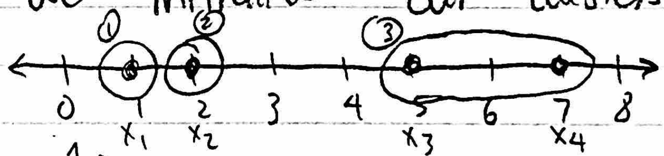
$$= (-\tfrac{1}{2})^2 + (\tfrac{1}{2})^2 + 0 + 0 \qquad \mu_3 = 7$$

$$= \tfrac{1}{4} + \tfrac{1}{4}$$

$$= \tfrac{1}{2}$$

$$\boxed{J(\{r_{nk}\}, \{\mu_k\}) = \tfrac{1}{2}}$$

b) Say we initialize our clusters like so:



\- Iteration 1:

0\. $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 6$

1\. $J(\{r_{nk}\}, \{\mu_k\})' = \sum_n \sum_k r_{nk} \|x_n - \mu_k\|_2^2$

$$= r_{11}\|x_1-\mu_1\|_2^2 + r_{12}\|x_1-\mu_2\|_2^2 + r_{13}\|x_1-\mu_3\|_2^2$$
$$+ r_{21}\|x_2-\mu_1\|_2^2 + r_{22}\|x_2-\mu_2\|_2^2 + r_{23}\|x_2-\mu_3\|_2^2$$
$$+ r_{31}\|x_3-\mu_1\|_2^2 + r_{32}\|x_3-\mu_2\|_2^2 + r_{33}\|x_3-\mu_3\|_2^2$$
$$+ r_{41}\|x_4-\mu_1\|_2^2 + r_{42}\|x_4-\mu_2\|_2^2 + r_{43}\|x_4-\mu_3\|_2^2$$

$$= 1\cdot 0^2 + 0 + 0$$
$$+ 0 + 1\cdot 0^2 + 0$$
$$+ 0 + 0 + 1\cdot(5-6)^2$$
$$+ 0 + 0 + 1\cdot(7-6)^2$$
$$= 1 + 1 = 2$$

2. $\mu_k = \dfrac{\sum r_{nk} x_n}{\sum r_{nk}}$

$\mu_1 = \dfrac{\sum_n r_{n1} x_n}{\sum_n r_{n1}} = \dfrac{1 \cdot 1}{1} = 1$

$\mu_2 = \dfrac{\sum_n r_{n2} x_n}{\sum r_{n2}} = \dfrac{1 \cdot 2}{1} = 2$

$\mu_3 = \dfrac{\sum_n r_{n3} x_n}{\sum_n r_{n3}} = \dfrac{1 \cdot 5 + 1 \cdot 7}{1+1} = \dfrac{12}{2} = 6$

3. repeat

Iteration 2:

0: $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 6$

1. $J(\{r_{nk}\}, \{\mu_k\}) = \sum_n \sum_k r_{nk} \|x_n - \mu_k\|_2^2$
$= 2$
(because means did not change)

2. $\forall k \in \{1, 2, 3\}$, μ_k remains the same

3. don't repeat b/c J does not change
$\Rightarrow$ Algorithm has converged

- This is suboptimal because there exists a lower J with a different clustering (part a)). Since this algorithm converged, and its J is still larger than that of a different clustering, it is suboptimal.

3. $\ell(\theta) = \sum_K \sum_n \gamma_{nK} \log \omega_K + \sum_K \{ \sum_n \gamma_{nK} \log(N(X_n | \mu_K, \Sigma_K)) \}$

a) $N(X | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$

$$N(X_n | \mu_K, \Sigma_K) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_K|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\frac{(X-\mu)^T(X-\mu)}{\Sigma_K}\right)}$$

$$\ell(\theta) = \sum_K \sum_n \gamma_{nK} \log \omega_K + \sum_K \{ \sum_n \gamma_{nK} \log\left(\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_K|^{\frac{1}{2}}} e^{-\frac{1}{2}((X-\mu)^T \Sigma_K^{-1} (X-\mu))} \right)$$

$$= \sum_K \sum_n \gamma_{nK} \log \omega_K + \sum_K \{ \sum_n \gamma_{nK} \left(\ln\left(\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_K|^{\frac{1}{2}}} \right) + \ln\left(e^{-\frac{1}{2}((X-\mu)^T \Sigma_K^{-1} (X-\mu))} \right) \right)$$

$$= \sum_K \sum_n \gamma_{nK} \log \omega_K + \sum_K \{ \sum_n \gamma_{nK} \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_K|^{\frac{1}{2}}} \right) + \gamma_{nK} \left(-\frac{1}{2} (X_n - \mu_K)^T \Sigma_K^{-1} (X_n - \mu_K) \right) \}$$

$$\frac{\partial}{\partial \mu_K} \ell(\theta) = 0 + \sum_n \gamma_{nK} \cdot -\frac{1}{2} \cdot (X_n - \mu_K)^T \cdot ((\Sigma_K^{-1})^T + (\Sigma_K^{-1}))$$

$$= \sum_n \gamma_{nK} \cdot -\frac{1}{\cancel{2}} \cdot (X_n - \mu_K)^T \cdot \cancel{2} \Sigma_K^{-1}$$ because Σ_K is symmetric

$$= \sum_n \gamma_{nK} \cdot (X_n - \mu_K)^T \cdot -\Sigma_K^{-1}$$

$$= \sum_n -\gamma_{nK} X_n \Sigma_K^{-1} + \gamma_{nK} \mu_K \Sigma_K^{-1}$$

$$\boxed{\frac{\partial}{\partial \mu_K} \ell(\theta) = \sum_n \gamma_{nK} \mu_K \Sigma_K^{-1} - \sum_n \gamma_{nK} X_n \Sigma_K^{-1}}$$

b) $\sum_n \gamma_{nK} \mu_K \Sigma_K^{-1} - \sum_n \gamma_{nK} X_n \Sigma_K^{-1} = 0$

$$\sum_n \gamma_{nK} \mu_K \Sigma_K^{-1} = \sum_n \gamma_{nK} X_n \Sigma_K^{-1}$$

$$\cancel{\Sigma_K^{-1}} \mu_K \sum_n \gamma_{nK} = \cancel{\Sigma_K^{-1}} \sum_n \gamma_{nK} X_n$$

$$\boxed{\mu_K = \frac{\sum_n \gamma_{nK} X_n}{\sum_n \gamma_{nK}}}$$

c) $K = 2$

n	1	2	3	4	5
x_n	5	15	25	30	40

n	γ_{n1}	γ_{n2}
1	0.2	0.8
2	0.2	0.8
3	0.8	0.2
4	0.9	0.1
5	0.9	0.1

$$w_1 = \frac{\sum_n \gamma_{nk}}{\sum_k \sum_n \gamma_{nk}} = \frac{(0.2+0.2+0.8+0.9+0.9)}{(0.2+0.2+0.8+0.9+0.9)+(0.8+0.8+0.2+0.1+0.1)}$$

$$= \frac{3}{3+2} = \frac{3}{5}$$

$$w_2 = \frac{(0.8+0.8+0.2+0.1+0.1)}{(0.2+0.2+0.8+0.9+0.9)+(0.8+0.8+0.2+0.1+0.1)}$$

$$= \frac{2}{5}$$

$$\mu_1 = \frac{1}{\sum_n \gamma_{nk}} \sum_n \gamma_{nk} x_n$$

$$= \frac{(0.2\times 5)+(0.2\times 15)+(0.8\times 25)+(0.9\times 30)+(0.9\times 40)}{(0.2+0.2+0.8+0.9+0.9)}$$

$$= \frac{1+3+20+27+36}{3} = \frac{87}{3} = 29$$

$$\mu_2 = \frac{(0.8\times 5)+(0.8\times 15)+(0.2\times 25)+(0.1\times 30)+(0.1\times 40)}{(0.8+0.8+0.2+0.1+0.1)}$$

$$= \frac{4+12+5+3+4}{2} = \frac{28}{2} = 14$$

$$\boxed{\begin{array}{ll} w_1 = \frac{3}{5}, & \mu_1 = 29 \\ w_2 = \frac{2}{5}, & \mu_2 = 14 \end{array}}$$